AWARD NUMBER: W81XWH-15-1-0243

TITLE: Development of a Prognostic Marker for Lung Cancer Using Analysis of Tumor Evolution

PRINCIPAL INVESTIGATOR: Edward F. Patz, Jr., M.D.

CONTRACTING ORGANIZATION: Duke University
Durham, NC

REPORT DATE: August 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**
August 2017	Annual	1 Aug 2016 - 31 Jul 2017

4. TITLE AND SUBTITLE Development of a Prognostic Marker for Lung Cancer Using Analysis of Tumor Evolution	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER W81XWH-15-1-0243
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Edward F. Patz, Jr., M.D. E-Mail: patz0002@mc.duke.edu	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Duke University Durham, NC 27705	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this project is to sequence the exomes of single tumor cells from tumors in order to construct evolutionary trees, the characteristics of which will be used to predict whether a tumor will metastasize or not. In the last project period, we have optimized tumor dissociation, tumor cell isolation, whole genome amplification, and exome sequencing. We have begun to sequence the exomes of single cells and to construct phylogenetic trees from the sequencing data. A sequencing pipeline has been established in order to obtain sufficient data with which to derive a prognostic classifier.

15. SUBJECT TERMS

NSCLC; tumor evolution; whole exome sequencing

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unclassified	6	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

1. Introduction

There is currently no consistent way to determine how aggressive or indolent a lung tumor will be even among patients with the same radiographic findings, histology, stage, or molecular markers. The purpose of this grant is to apply evolutionary analytical methods developed to study expansion and migration of populations to tumor biology in order to produce a prognostic marker in cancer. As with the Darwinian evolution of populations, the evolution of tumor cells within a tumor can be diagrammed on a phylogenetic tree. The more diverse a tumor's phylogenetic tree, the more likely it is that there are cells within it that have acquired the genetic alterations that allow them to proliferate at an increased rate, migrate, and metastasize. We will develop and validate a novel, objective, and measurable "prognostic score" based on the probability that some tumors will be *aggressive* and metastasize, and other tumors will be *indolent* and not metastasize. We first will perform whole exome sequencing of individual tumor cells from the tumors of a training set of patients (half early stage, half late stage). We will reconstruct each tumor's phylogenetic tree (a map of the clonal evolution reflecting divergence and heterogeneity), and compare the tree patterns from early stage NSCLC (indolent tumors without metastasis) to those from late stage disease (tumors with metastasis). We will use a combination of *tree features* (including branch length and tree shape) to generate a prognostic score (a continuous variable and a measure of tumor heterogeneity) that separates tumors with very different phenotypes (indolent vs. aggressive). We will derive the prognostic score by determining the probability of each individual tumor's outcome in the pilot training study, and then validate this strategy in an independent set of patients. An accurate prognostic score could significantly change clinical management and improve outcomes.

2. Keywords

NSCLC; tumor evolution; whole exome sequencing

3. Accomplishments

Specific Aim I: Isolate individual tumor cells from 10 patients with stage I non-recurrent NSCLC and 10 patients with advanced stage NSCLC.

The previous report described our efforts in optimizing the technical aspects of tumor dissociation, tumor cell isolation, and whole genome amplification (WGA) in preparation for exome sequencing. All technical issues have now been satisfactorily resolved. We have enrolled a total of 24 patients so far and isolated their tumor cells, although the yield has been variable. We continue to enroll patients and isolate tumor cells.

Specific Aim II. Perform single cell whole exome sequencing on 40 individual cells isolated from each tumor.

We have established a standard operating procedure for obtaining whole exome sequence from the amplified DNA from single tumor cells. DNA is subjected to standard next generation sequencing library preparation and exome capture. Briefly, library preparation is performed using the KAPA Biosystems HyperPlus kit (Roche) according to the manufacturer's protocol. This kit includes enzymatic fragmentation, end repair, A-tailing, and barcode adapter ligation to create sequencing libraries. After the libraries are made, they are subjected to exome capture using the IDT Research Exome (Integrated DNA Technologies) kit according to the manufacturer's instructions. This process includes hybridization and pulldown of target sequences. Finally, 96 barcoded samples are sequenced on a single Illumina HiSeq 2500 high output flow cell. Sequencing is performed using paired end sequencing with a read length of 125 base pairs. Basic quality control is performed before the data is subjected to bioinformatic analysis.

Using this standard procedure and as described below, we compared 6 plex, 12 plex and 24 plex sequencing and found 12 plex sequencing to give the best overall quality. We have now instituted a sequencing pipeline. Phylogenetic analysis of the tumor cells of the first patient has begun, as described in the next section.

Specific Aim III. Using the whole exome sequence data, analyze phylogenetic relationships of tumor cells and develop a prognostic classifier.

Mutation calling

We performed single cell sequencing data for 12 single cells from one patient and 1 single cell cell-line control in 6 plex, 12 plex and 24 plex. We found the 12 plex sequencing rounds have the best overall quality with coverage of 26, mapping ratio of 0.998 and duplication ratio of 0.48. Therefore, we used 12 plex sequencing data and 1 single cell cell-line control for mutation calling.

Raw reads were aligned to reference genome GRCh37 by BWA MEM algorithm [1]. We followed the GATK [2] best practice guide [3] to mark duplicated reads, recalibrate quality score, call variant separately for each sample with HaplotypeCaller in GVCF mode and jointly genotype all samples together.

Phylogenetic analysis

For phylogenetic analysis, we selected 21,285 variant sites with no missing information for all samples (read coverage>0). These sites were concatenated to one sequence for each sample and heterogenetic sites were represented with standard ambiguity code (IUPAC). We used Smart Model Selection (SMS)[4] with Akaike Information Criterion to select the Generalized Time Reversible [5] with gamma distributed rate (GTR+G) model as the best fitted nucleotide substitution model for our data. Then we used PhyML [6] to build a Maxim-likelihood tree with this model. Approximate Likelihood-Ratio Test (aLRT) [7] was used to provide the branch support. The tree was rooted on the cell-line control sample.

The phylogenetic tree (**Figure 1**) shows generally good support (aLRT >0.75) except for 3 internal nodes. There are three main clades on our tree, which may suggest three sub-clonal populations and they are in a sequential developing pattern. More cells from this patient's tumor will be analyzed to verify this evolutionary pattern.

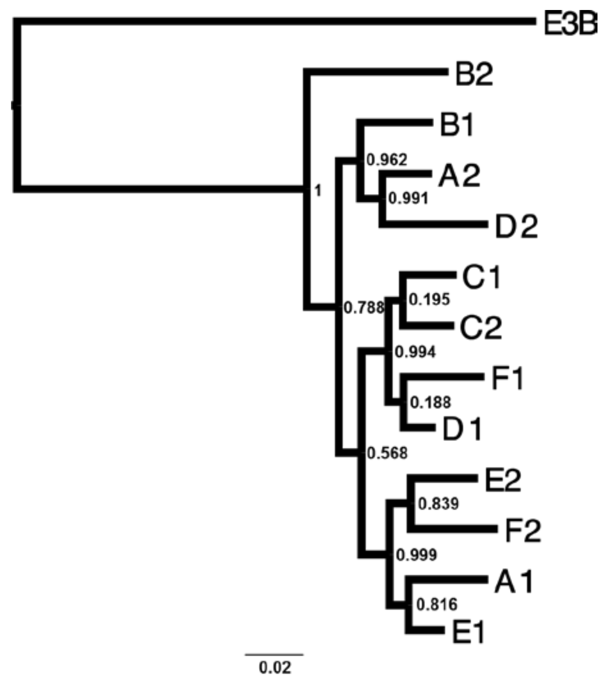

Figure 1. Preliminary phylogenetic tree.

As more sequence data is obtained, we will continue to analyze phylogenetic relationships of tumor cells in the tumors of all 20 patients and develop a prognostic classifier.

Specific Aim IV. Validate the prognostic classifier developed in Specific Aim III in an independent blinded study.

Nothing to Report.

Opportunities for training and professional development

Nothing to Report.

How results were disseminated to communities of interest

Nothing to Report.

Plans for next reporting period

In the next reporting period, we plan to finish all of the sequencing for this project and develop a prognostic classifier, completing Specific Aim III. We will then prepare to begin Specific Aim IV to validate the prognostic classifier.

4. Impact

Impact of the development of the principal discipline of the project
Nothing to Report

Impact on other disciplines
Nothing to Report

Impact on technology transfer
Nothing to Report

Impact on society beyond science and technology
Nothing to Report

5. Changes/Problems

Nothing to Report

6. Products

None to date

7. Participants & Other Collaborating Organizations

E.B. Gottlin, investigator, performed tumor cell isolation, 2.4 cal. months
S.G. Gregory, investigator, supervised WGA, 0.48 cal. months
E.F. Patz, Jr., PI, 1.36 cal. months
EA Burns, Lab Assistant, 2.4 cal. months
A.G. Rodrigo, 0.15 cal. months
Y. Ding, Graduate Assistant, 0.15 cal. months

We have begun collaborating with LabCorp, who will be providing some of the DNA sequencing; there has been no change in the active support of the PI.

8. Special Reporting Requirements

None

9. Appendices

None